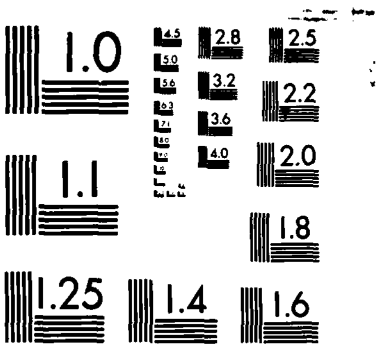
AD-A157 583 THE TRACKING OF REFERENTS IN DISCOURSE: AUTOMATED 1/1
VERSUS ATTENDED PROCESSES(U) OREGON UNIV EUGENE DEPT OF
PSYCHOLOGY T GIVON ET AL. 15 MAY 85 TR-85-4-ONR
UNCLASSIFIED N00014-83-K-0601 F/G 5/10 NL

END
FILMED
DTIC


1.0
4.5
5.0
5.6
6.3
7.1
8.0
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(2)

AD-A157 583

# Cognitive Science Program

## THE TRACKING OF REFERENTS
## IN DISCOURSE:
## AUTOMATED VS. ATTENDED PROCESSES

by

T. Givón, W. Kellogg, M. Posner
and P. Yee

Technical Report 85-4

## University of Oregon

DTIC
S ELECTE D
JUL 1 7 1985
G

DTIC FILE COPY

Research Sponsored by:

Personnel and Training Research Programs
Psychological Services Division
Office of Naval Research
Under Contract No. N00014-83-K-0601
ONR No. 85-4

Reproduction in whole or in part is permitted
for any purpose of the United States Government

Approved for public release: distribution unlimited

85 06 25 256

# DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS<br>BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>ONR No. 85-4 | 2. GOVT ACCESSION NO.<br>AD-A157 583 | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>The Tracking of Referents in Discourse: Automated vs. Attended Processes | | 5. TYPE OF REPORT & PERIOD COVERED<br>Final Report |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>T. Givón, W. Kellogg, M. Posner, P. Yee | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-83-K-0601 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS | | 12. REPORT DATE<br>May 15, 1985 |
| | | 13. NUMBER OF PAGES<br>42 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>Unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

psycholinguistics, discourse processing, referential coherence, anaphor resolution, syntax of reference, cognitive processes in coreference, attentional vs. automatic processing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The cognitive processes involved in achieving the overall coherence of discourse are central to an understanding of the use of natural language in communication. The present series of studies examines the mental effort involved in achieving coreference. We explore a major factor that affects the accessibility of referents: the length of absence of the referent from the distribution of different types of grammatical referential devices in discourse have shown a correlation between the type of device (e.g. full noun)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102 LF 014 6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DTIC
COPY

| Accession For | |
|---|---|
| NTIS GRA&I | ☒ |
| DTIC TAB | ☐ |
| Unannounced | ☐ |
| Justification | |
| By | |
| Distribution/ | |
| Availability Codes | |
| Dist | Avail and/or Special |
| A-1 | |

# THE TRACKING OF REFERENTS IN DISCOURSE: AUTOMATED VS. ATTENDED PROCESSES

T. Givón, Linguistics Department
W. Kellogg, Psychology Department
M. Posner, Psychology Department
P. Yee, Psychology Department
and Cognitive Science Program
University of Oregon, Eugene

## 1. Introduction

### 1.1. Referential coherence

The processes involved in creating the overall coherence of human discourse are central to our understanding of the use of natural language in communication. Of the many factors that contribute to overall coherence, the most concrete, obvious and well-studied is **referential coherence.** It involves the introduction, re-introduction, identification and tracking of referents in discourse, primarily nominal referents that most commonly fill the subject or object roles in clauses. While it is possible to study referential coherence in isolation, the effect of other components of discourse coherence on the processing of referents (see e.g. Anderson, Garrod & Sanford, 1983) also must be considered.

### 1.2. Grammatical devices used in anaphoric reference

In a recent series of quantified text studies involving many languages (Derbyshire, 1985; Givón, 1983(a); Payne, 1985; Tomlin, 1985; **inter alia**) it was shown that the grammatical devices used for referent identification in discourse can be scaled according to the **predictability**, **accessibility** or **continuity** of reference. The accessibility of referents in discourse is affected by four major factors, only two of which are directly associated with reference:

(1) **Factors affecting referential predictability:**

(i) Length of absence of the referent from an active discourse file, an effect that is presumably due to some facet of memory decay;
(ii) Potential interference from other referents within the immediately preceding discourse environment;
(iii) Semantic information from inside the clause, in terms of semantic-selectional restrictions, which may contribute to disambiguation;
(iv) Thematic information from higher organizational levels of the discourse, which may also contribute to disambiguation.

In recent discourse studies (see Givon, 1983a), two text-based anaphoric ("backward-looking") measures of referential accessibility were developed. They correlate rather transparently with the ease of assigning coreference as affected by factors (i) and (ii) above. Further, they suggest the possibility that text-based measures of referential continuity may correlate with experimentally measured mental capacities, presumably the mental effort used in assigning coreference. Those two text-based measures are:

(a) **Referential distance:** Measured from the last occurrence in the preceding discourse (in number of clauses to the left);
(b) **Referential complexity:** An assessment of the number of semantically compatible nominals in the immediately preceding discourse (3 clauses).

The results of both measures rank-order the most common anaphoric coreference devices found in grammars as follows:

(2) **most predictable/accessible/continuous coreference**
a. zero anaphora
b. anaphoric unstressed pronouns
c. contrastive/stressed pronouns
d. definite nouns
e. restrictively-modified definite nouns
**least predictable/accessible/continuous coreference**

More specifically, **anaphoric pronouns** were shown to be used in discourse with an average referential distance of **1.0 clauses** and relatively **low** referential complexity of the directly preceding environment. In contrast, **definite nouns** tended to exhibit an average referential distance of **10.0 clauses** and relatively **high** referential complexity of the directly preceding environment. These distributional patterns suggest that anaphoric pronouns may be preferred for signalling coreference over short referential gaps and for low referential complexity. On the other hand, full definite nouns may be more efficient anaphoric devices for signalling coreference over long referential gaps and for high referential complexity. Further, Tomlin (1985) has shown that definite nouns are favored over pronouns at points of major **thematic discontinuity** in discourse. In contrast, anaphoric pronouns are favored over definite nouns when thematic continuity is high. One may thus wish to conclude, tentatively, that anaphoric pronouns are more efficient referential devices in discourse contexts of **high informational predictability**, while definite nouns are more efficient devices in discourse contexts of **low informational predictability**.

### 1.3. Grammar as an automated language-processing device

Grammatical devices -- such as rigid word order or morphology -- may be viewed as routinized, or automated, strategies that speed up the processing of linguistically coded information. There exists, in fact, a pre-grammatical mode of language processing, used in early childhood or early second-language acquisition (Pidginization). This less routinized processing mode can be compared with the fully grammaticized native/adult language along the following socio-functional dimensions (see Givón, 1979, ch. 5; Givón, 1982):

(3)

| | PRE-GRAMMATICIZED MODE (unroutinized) | GRAMMATICIZED MODE (routinized) |
|---|---|---|
| a. | Slower processing speed | Faster processing speed |
| b. | Higher uncertainty/ambiguity of the communicative context | Lower uncertainty/ambiguity of the communicative context |
| c. | A society of intimates | A society of strangers |
| d. | Greater shared background | Smaller shared background |
| e. | Low socio-informational complexity | High socio-informational complexity |
| f. | Greater reliance on context | Lesser reliance on context |

In assigning coreference in discourse processing, one would obviously expect factors such as those listed in (1) above to exert their influence somewhat independently of grammar. Nonetheless, if our hypothesis about grammar as an automated/routinized processing mode is correct, one ought to be able to measure the effect of grammar itself during referential processing, when the factors listed in (1) are held constant. The series of experiments reported here purport to measure such effects. In the experimental psychology literature, the difference between attended and unattended ('automated', 'routinized') processing has been studied at many levels of cognition, neurology and motor-control (e.g. Keele, 1968; Posner & Marin, 1985; Posner & Warren, 1972; Schneider & Shiffrin, 1977; Shapiro & Schmidt, 1982; Whitaker, 1983). Our study thus purports to assess the independent effect of grammar on the processing of coreference.

## 2. Methodological considerations

### 2.1. The measurement of mental processes

In attempting to understand how referential coherence is achieved, it is important to devise ways for measuring the covert mental processes that take place during the introduction and re-introduction of referents in actual text processing. There are also important theoretical reasons in both psychology and linguistics for developing such measurements. Two basic methods have been used in the past to study covert attentional processes:

(a) Measuring the **time** required to perform a certain task;
(b) Measuring the **capacity** required to perform various tasks.

In order to assay the time spent on the processing of text, one may measure the rate of reading, the duration of fixation on a single word, or the time required for the keypress release of successive words. The last method has a certain advantage in yielding a word-by-word measurement in reading. It may perhaps also have certain disadvantages, in that it is not the natural mode of reading. Existing evidence suggests that at least under instructions to retain text information verbatim this measure is sensitive to different mental processes (Aaronson & Ferres, 1983; Just and Carpenter, 1980). However, under conditions calling for comprehension of gist, the keypress method may be less sensitive (Aaronson & Ferres, 1983).

Another type of method assays covert mental processes by inserting a secondary task in which the subject must respond to a relatively rare **probe event** by pressing a special key. While this method has been used in many studies (see Kerr, 1973, for a review), its use has been confined in the past to assessing the overall difficulty of text passages, but not the mental effort required for the processing of individual lexical items. In this work, we have combined both methods in order to obtain a more detailed picture of the mental processes involved in the establishment of coreference during the reading of text.

Finally, it is also possible to measure the efficiency of various grammatical devices used in referential processing by interrupting the reading task and requiring the subject to make a decision that reflects the achievement of coreference. In one of the present experiments, we asked subjects to make such a decision immediately after the occurrence of anaphoric grammatical devices of different kinds. Such a measure can be used as an independent control, in addition to its intrinsic value.

In the experiments reported below, we measured the key press time for the release of words in continuous reading of text passages. The key press times for the release of referential words -- pronouns or definite nouns -- were measured in the presence of a visual probe; detecting this visual probe thus constituted a secondary, attention-demanding task, involving a separate key press from which probe reaction time could be assessed independently.

## 2.2. Text frames

The discourse frames used involved highly conventional episodes or scripts, all involving familiar culturally-stereotyped scenes, such as "at the bank", "at the store", "the piano lesson", "after the theater", "at a restaurant", etc.. In all episodes, two participants were involved, one male, the other female. Sex-roles were conventionally stereotyped, and the gender identity of both participants was explicitly indicated. This was done so that when the pronouns 'he' or 'she' were used, no ambiguity of reference would arise. The role of both participants in the "scene" was highly stereotyped, given general cultural conventions associated with each scene/episode. This was done to insure that the thematic frame itself would exert as little influence as possible toward referential identification of either of the two participants. Characteristic sample texts are given below for each experiment.

## 3. Experimental method and results

### 3.1. Experiment 1

#### 3.1.1. Method

#### 3.1.1.1. Materials and design

In this experiment there were ten short story themes that introduced two characters (always one male and one female) and ten that introduced a single character. Six different versions of these stories were created by crossing two factors: referential distance (0 or 5 clauses between subsequent references to a character) and anaphoric device (the type of anaphor used: an identical noun, a pronoun or an epithet). All of the single character stories and the two epithet versions of the dual character stories served as filler material to increase the variety of materials presented to subjects and to reduce expectancy effects. All of the experimental paragraphs were two-character stories and introduced the characters within the first two sentences of the story. The next five clauses continued to discuss only one of the two characters introduced. In one type of experimental paragraph, the last sentence used either a pronoun or a noun that was identical to the one used in the first sentence to refer to the main character -- the one mentioned in the intervening clauses. This created a story with a one-clause referential distance between the last mention of the referent and the subsequent use of the anaphor. In the other type of experimental paragraph, the anaphor in the last sentence referred to the secondary character of the story -- the one not mentioned in the intervening clauses. This variation created a story with a five-clause referential distance.

Other versions of the two-character stories served as filler material and differed only in the type of anaphor used in the last sentence. As in the experimental paragraphs, the anaphors might refer to either of the two characters, but instead of using an identical noun or a pronoun to refer to the characters these filler paragraphs used other nouns such as 'man' or 'woman.' These were called "epithets" in this study. All of the single character stories were filler paragraphs, and their form followed a pattern similar to the experimental paragraphs. The first sentence introduced only one character, and the subsequent sentences discussed material unrelated to the character introduced. The last sentence then referred back to the character using either a pronoun, an identical noun or an epithet. Examples of the stimuli are presented in (4) below.

(4) **Sample text passage for Experiment 1:**

a. **Two character frame**

THE STAR rushed to the newsstand to get first crack at the review of the show. HER MANAGER rushed right behind her. He reached the newsstand first, grabbed a paper, tore it open, scanned it and stopped abruptly.

1-CLAUSE ENDINGS:

(i) Identical Noun: The *manager...
(ii) Pronoun: *He...
(iii) Epithet (filler): The *man...
...said in a loud voice: "I told you so, didn't I?"

5-CLAUSE ENDINGS:

(i) Identical Noun: The *star...
(ii) Pronoun: *She...
(iii) Epithet (filler): The *woman...
...said in a loud voice: "I told you so, didn't I?"

b. **Single character frame**

THE FARMER drove an old Ford pickup to haul his hay. It was a classic Model-T with bent fenders and a rusted top, but the engine was in top shape and purring smoothly even under a heavy load of fifty bales. Near the barn gate...

ENDINGS:

(i) Identical Noun: ...the *farmer...
(ii) Pronoun: ...*he...
(iii) Epithet (filler): ...the *man...
...stopped the truck and stepped down.

The different versions of the twenty paragrphs were distributed among three stimulus lists such that across subjects all of the paragraphs appeared equally often with identical noun, pronoun and epithet anaphors. Each list contained two of the six versions of each story frame, one at each referential distance, such that an equal number of one- and five-clause-distance paragraphs occurred in each list. Subjects were tested with only one list, so within a testing session subjects saw two versions of all paragraphs. Within each of the paragraphs were preselected probe locations. Probes always occurred simultaneously with the anaphoric devices. Half of the paragraphs had additional, pseudorandomly chosen probe sites, which were never within ten words of the experimentally probed locations. Control probes appeared in the same locations for subjects tested with the same list. Seven practice paragraphs were constructed similar to the paragraphs in the experiment to familiarize subjects with the task. True/false comprehension questions for each story in the practice trials and in the experimental trials were prepared in a small testing booklet.

In sum, the design of this experiment was a 2x2 factorial, with referential distance (0 or 5 clauses) and type of anaphor (identical noun vs. pronoun) as repeated measures.

### 3.1.1.2. Subjects

Twelve individuals recruited from the subject pool of the Cognitive Lab at the Psychology Department, University of Oregon, served as subjects for this experiment. All were native speakers of English, had normal or corrected-to-normal vision, and had no apparent reading disabilities. Subjects were paid $4.00 for participating in the one hour experiment.

### 3.1.1.3. Procedure

Subjects completed the practice trials to familiarize themselves with the experimental procedure. Each subject was tested individually with one of the stimulus lists. At the start of each trial a 'READY' message appeared in the lower left corner of the computer monitor. To begin a trial, subjects pressed the space bar on the computer keyboard. This started presentation of a paragraph. Paragraphs appeared one word at a time beginning in the upper left corner of the screen. The words were presented as though a window opened and closed across the page to reveal successive words and cover previous ones. Subjects controlled the presentation rate by pressing the space bar for each word. This method is similar to the sequential condition used by Kennedy and Murray (1984).

Subjects were instructed to read the paragraphs at their normal reading rate and to understand the stories because their comprehension would be tested after each trial. Subjects were also informed of the possible occurrence of probes (a flashing square of light) that would appear at the top center of the screen. Whenever a probe occurred, subjects were to press a key as quickly as possible with their right hand while continuing to press for words with their left. Thus, while subjects pressed the space bar with their left hand to read the stories, they rested their right hand on a separate key to press whenever a probe occurred. As mentioned previously, the probed locations were predetermined for all stories. When subjects responded to a probe, it disappeared from the screen. When subjects reached the end of a paragraph the computer instructed them to answer a comprehension question that corresponded to the story just presented. Subjects were given as much time as they wanted to answer the comprehension questions. When they had finished marking their answer, they pressed the 'RETURN' key on the computer keyboard and received the 'READY' message for the next trial.

Stimulus presentation and data collection were controlled by an Apple II Plus computer system. All text was presented in uppercase letters. Paragraphs were presented in a different random order for each subject. Reaction times to probes and reading times for each word (inter-keypress intervals) were recorded.

## 3.1.2. Results

### 3.1.2.1. Comprehension test

The subjects' mean percent correct on the comprehension test was 84%. It thus appears that the subjects were reading and understanding the stories as instructed.

### 3.1.2.2. Probe reaction time

The means of the subjects' median reaction times to the visual probes in each condition were calculated and are given in Table I, below, and represented graphically in Figure I, below. Reaction times are expressed in milliseconds.

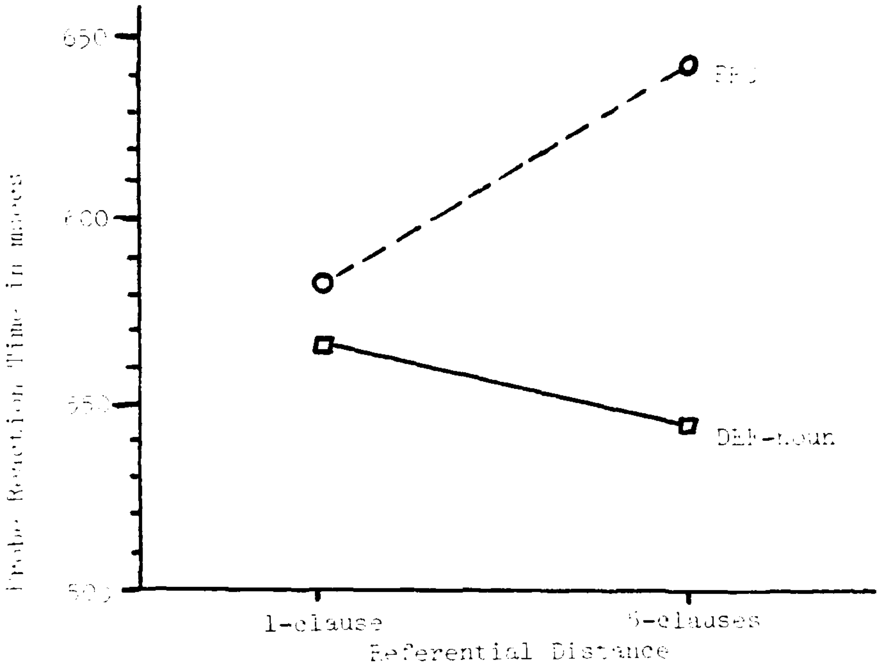
TABLE I: Median Probe Reaction Time (in msec.) for
Pronouns and Anaphoric Definite Nouns at
1 and 5 Clause Referential Distance

| | REFERENTIAL DISTANCE | |
|---|---|---|
| | 1-clause | 5-clause |
| PRONOUN | 584 | 644 |
| DEF-NOUN | 567 | 547 |

(CONTROL PROBE: 620)

In these results, reaction times (RTs) for probed pronouns are longer than for the probed DEF-nouns. In terms of sensitivity to changes in referential distance, the probe RT for pronouns seems to increase with the increase in referential distance from 1 to 5 clauses, while the RT for DEF-nouns seems to decrease with the increase in referential distance. Statistical analysis, however, indicated that none of the effects (referential distance, type of anaphor or their interaction) is significant ($F(1,11) = .89$, 2.89 and 1.92 for the three effects, respectively).

FIGURE I: Median Probe Reaction Time for Pronouns and
Anaphoric Definite Nouns at 1 and 5 Clause
Referential Distance



### 3.1.2.3. Keypress times for probed words

Subjects' mean inter-keypress times for pronouns and DEF-nouns at the 1 and 5 clause referential distances are given in Table II and expressed graphically in Figure II below. We list the keypress time for the probed referential device (n), as well as for the two words preceding and following it.

The most striking result here is the difference in the peaks of the keypress times for the two referential devices. For pronouns, the longest press time is on the probed pronoun itself [n]. In contrast, for DEF-nouns the longest press time is on the word following the probed noun [n+1]. This interaction between type of anaphor and word position (n-2, n-1, n, n+1, n+2) was significant ($F(10,110) = 2.62$; $p < 0.01$). Statistical analysis also revealed a main effect of word position due to the increased times around the probed items ($F(10,110) = 4.64$; $p < 0.01$).

The experiment revealed no significant effect of increased referential distance on the keypress times for either pronouns or DEF-nouns. Thus there was no interaction between referential distance and word position, nor between referential distance, type of anaphor and word position (both $F < 1$).

TABLE VII: Median Probe Reaction Time for Anaphoric-DEF and Thematic-DEF Nouns at Different Referential Distances

| CONDITION | REFERENTIAL DISTANCE<br>1-CLAUSE | 10-CLAUSE | 20-CLAUSE |
|---|---|---|---|
| anaphoric-DEF | 710 | 674 | 697 |
| thematic-DEF | 734 | 698 | 736 |
| control probe | 713 | 677 | 688 |

Statistical analysis reveals a significant main effect of type of referential device, with anaphoric DEF-nouns and controls showing faster probe RTs than thematic DEF-nouns (*F*(1,35) = 7.35, p<.02). The effect of referential distance is also significant, with 10-clause distance RTs being faster that either 1-clause or 20-clause (*F*(2,70) = 3.23, p<.05), similar to the U-shaped pattern for pronoun probe RTs observed in Experiment 3. There is no significant interaction between type of referential device and referential distance (*F*<1).

### 3.4.2.3. Keypress times for probed words

The results of the keypress times for the various conditions are given in Table VIII, below.

TABLE VIII: Mean Keypress Times for Anaphoric-DEF and Thematic-DEF nouns at Different Referential Distances

| CONDITION | WORD POSITION<br>n-2 | n-1 | n | n+1 | n+2 |
|---|---|---|---|---|---|
| ANAPHORIC-DEF: | | | | | |
| 1-clause | 296 | 302 | 456 | 509 | 319 |
| 10-clause | 297 | 306 | 426 | 481 | 325 |
| 20-clause | 343 | 441 | 458 | 470 | 306 |
| THEMATIC-DEF: | | | | | |
| 1-clause | 323 | 411 | 513 | 492 | 483 |
| 10-clause | 314 | 326 | 444 | 462 | 333 |
| 20-clause | 356 | 320 | 566 | 467 | 321 |
| CONTROL PROBE: | | | | | |
| 1-clause | 319 | 320 | 545 | 518 | 378 |
| 10-clause | 330 | 309 | 486 | 485 | 324 |
| 20-clause | 301 | 304 | 448 | 501 | 295 |

For anaphoric-DEF nouns, the peak keypress time again appears on the word following the probed one (on n+1). Further, the keypress time at the [n+1] position grows shorter with the increase in referential distance from 1 to 20 clauses. This effect is represented graphically in Figure VI. The decrease in press times for

Subjects were randomly assigned to one of three groups, with each group tested on only one of the lists. Half of the subjects in a group were tested with one block first, the other half with the other block first. Hence, the design of this experiment was a 2x3 factorial, with referential device (anaphoric-DEF, thematic-DEF) and referential distance (1,10,20 clauses) as repeated measures.

As before, probes were presented with the nouns of the critical noun phrases. Control probes were presented at pseudorandom locations within some of the passages. The number of control probes varied from 0 in the shortest passages to 3 in the longest. True/false questions were prepared and administered to the subjects to test their comprehension of the stories.

#### 3.4.1.2. Subjects

Thirty-six subjects were recruited from the same pool as described above, subject to the same conditions outlined above.

#### 3.4.1.3. Procedure

The same procedure was used here as in Experiment 1, except that the subjects pressed the "?" key on the computer keyboard to respond to probes. Stimulus presentation and data collection were controlled by an Apple IIe computer system.

### 3.4.2. Results

#### 3.4.2.1. Comprehension test

The subjects' mean percent correct on the comprehension test was 91%, again demonstrating that they followed instructions and understood the stories.

#### 3.4.2.2. Probe reaction time

Median probe reaction times for the experimental conditions as well as for control items are given in Table VII, below. The experimental conditions show a weak U-shaped pattern, with a dip in the RTs for the 10-clause distance. This is essentially the control probe pattern, to which the anaphoric-DEF noun pattern conforms rather closely. The pattern is similar though a bit more pronounced for the thematic-DEF nouns.

## 3.4. Experiment 4

### 3.4.1. Method

#### 3.4.1.1. Materials and design

This experiment compared the behavior of definite nouns -- either anaphoric or thematic -- at three referential distances: 1 clause, 10 clauses and 20 clauses. As in Experiment 2, the test passages introduced one character at the beginning. Intervening material of varying length then dealt with the setting and related matters, but avoided mentioning the character. Finally, the ending either mentioned the character again (anaphoric-DEF noun), or alternatively introduced a second character that was highly stereotyped given the particular scene, and thus thematically predictable (thematic-DEF noun). A sample story frame can be seen in (7) below.

(7) **Sample test passage for Experiment 4**

1-CLAUSE DISTANCE OPENING:
The woman waited patiently near the counter. The shop was almost empty and nobody seemed to notice her...

10-CLAUSE DISTANCE ADDITION: It was a large store with many departments spread along the cavernous floor space of three main levels. The time was early in the afternoon on a Monday in late spring, and nobody seemed to be moving around. On weekends the store was always packed with hoards of bargain shoppers milling in the aisles, holding sale coupons in hand, and craning their necks in search of that special deal that was somehow always two counters away, or so it seemed...

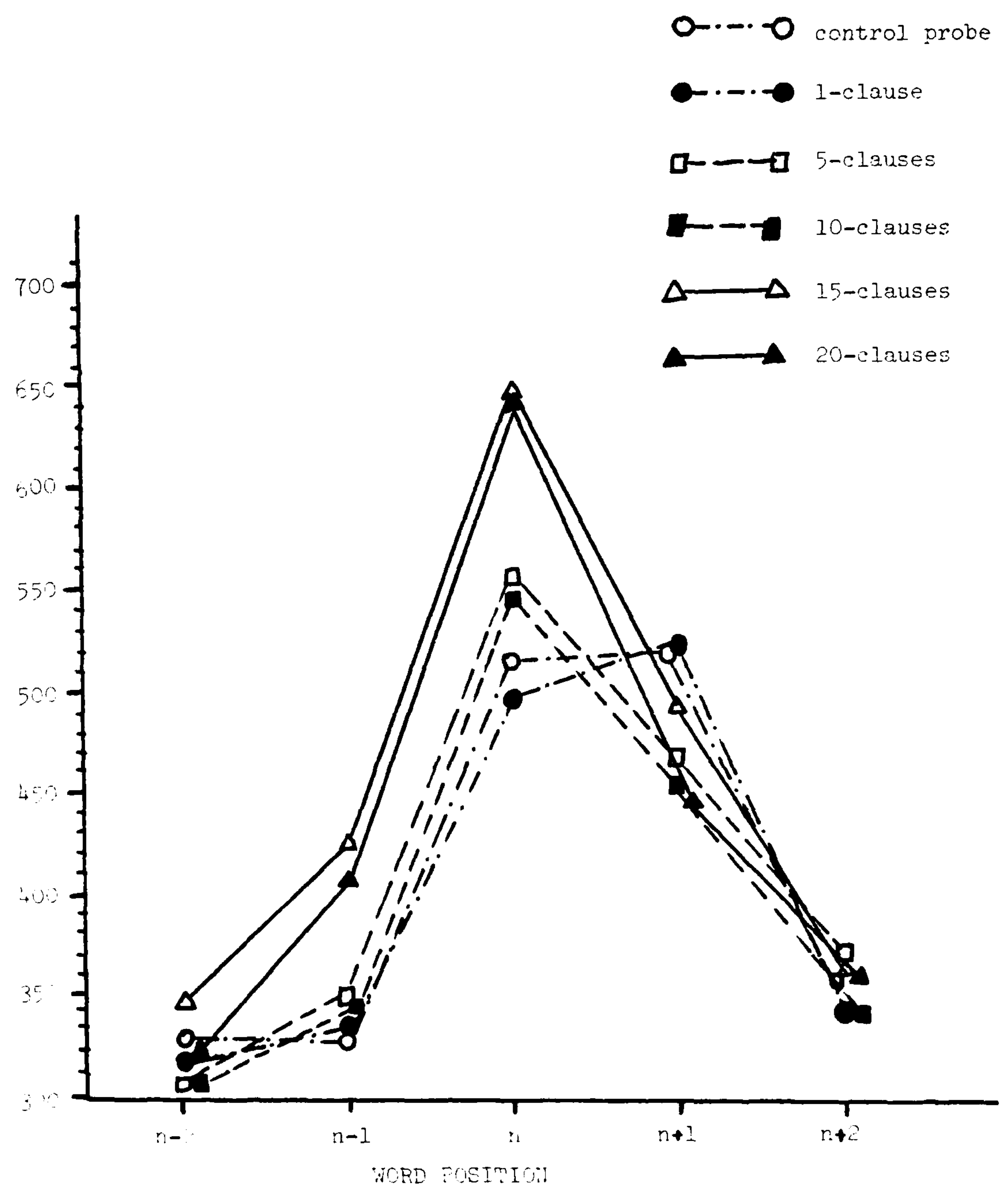
20-CLAUSE DISTANCE ADDITION:
The store was part of a chain that sprang up, seemingly instantly, in almost every shopping center of sufficient size in the burgeoning suburbs of big towns. It specialized in high-volume, medium-priced merchandise. The same national company always seemed to submit the successful bid. With standardized merchandise and synchronized sales, it was a shopper's paradise wherever it was located...

ANAPHORIC-DEFINITE ENDING:
The *woman took out her compact and touched up her hair.

THEMATIC-DEFINITE ENDING:
The *salesman finally came over and asked her what she was looking for.

There were twelve test stories, each with six versions. Subjects read two versions of each story frame, and in addition six naturalistic filler passages taken from a current novel. There were three stimulus lists consisting of an equal number of paragraphs in each condition. The lists were divided into two blocks, each block containing only one version of a test story. Across subjects, the stories appeared an equal number of times across in each condition.

FIGURE V: Mean Keypress Times for Pronouns
at Different Referential Distances



If one accepts the explanation of the inordinately high probe RT for the 1-clause distance, then the results of the probe RT measure here replicate and extend the trend observed in Experiment 1 (see Figure I).

### 3.3.2.3. Keypress times for probed words

The keypress times for this experiment are given in Table VI and represented graphically in Figure V, below.

TABLE VI: Mean Keypress Times for Pronouns
at Different Referential Distances

| REFERENTIAL DISTANCE | WORD POSITION n-2 | n-1 | n | n+1 | n+2 |
|---|---|---|---|---|---|
| 1-cl | 316 | 335 | 499 | 524 | 340 |
| 5-cl | 305 | 350 | 558 | 470 | 373 |
| 10-cl | 304 | 339 | 547 | 456 | 341 |
| 15-cl | 348 | 425 | 650 | 491 | 366 |
| 20-cl | 320 | 407 | 643 | 443 | 361 |
| control probe | 328 | 323 | 519 | 528 | 357 |

The pattern of peak keypress times found for pronouns in this experiment is similar to the one found for pronouns in Experiment 1. That is, the longest press time is for the probed word (n). Also as in Experiment 1, the longest press time for control items is for the word following the probed one (n+1). More importantly, a clear effect of referential distance on processing time is discernible: increases in referential distance produce increases in the peak keypress time for the probed item. At the short 1-clause distance the increase in keypress time at (n) is not as great as for the 5 and 10-clause distances, which in turn is not as great as the increase for the 15 and 20-clause distances. These groupings conform rather closely to the probe reaction times reported above (with the exception of the 1-cl distance, discussed there).

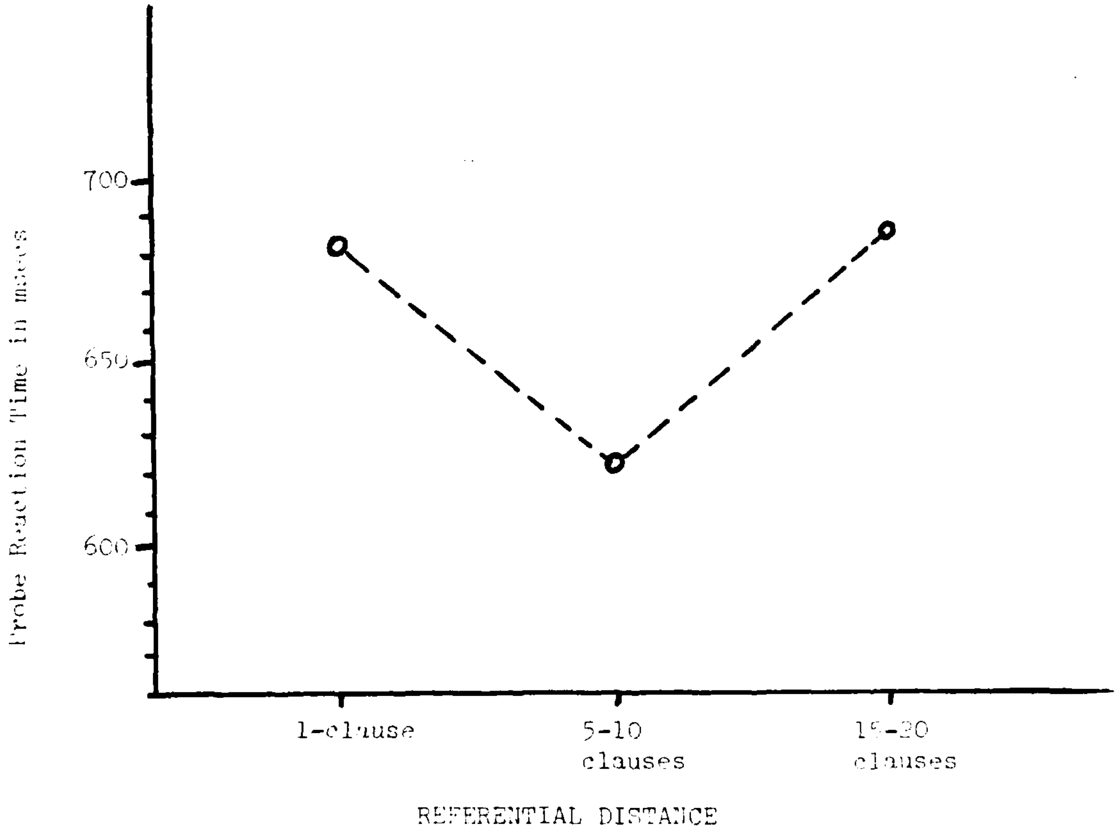
The statistical analysis shows that the main effects are both significant: for referential distance ($F$ (5,10) = 7.25; $p < 0.001$), and for word position ($F$ (10,20) = 49.19; $p < 0.001$). The interaction between referential distance and word position is also significant ($F$ (50,100) = 3.90; $p < 0.001$).

TABLE V: Median Probe Reaction Time for Pronouns at Different Referential Distances

| CONDITION | RT |
|---|---|
| 1-clause | 683 |
| 5-clause | 626 |
| 10-clause | 617 |
| 15-clause | 684 |
| 20-clause | 691 |
| Control probes | 654 |

The probe reaction times are presented graphically in Figure IV, below.

FIGURE IV: Median Probe Reaction Time for Pronouns at Different Referential Distances



for each paragraph in a stimulus file, was prepared and administered to the subjects in order to make sure that they read and understood the stories.

### 3.3.1.2. Subjects

Thirty-six individuals who had not participated in any previous experiment were recruited from the same subject pool as above, with the same conditions applying (see above).

### 3.3.1.3. Procedure

The same procedure was followed as in the previous experiment, except that the subjects did not use the space bar of the computer to press for the presentation of words. Instead they used a key located to the left of the computer. As before, they used a separate key with their right hand to respond to the probes.

### 3.3.2. Results

### 3.3.2.1. Comprehension test

The subjects' mean percent correct on the comprehension test was 91%.

### 3.3.2.2. Probe reaction time

The means of the subjects' median probe reaction times were calculated and are given in Table V, below. A "U-shaped" pattern is evident, with slow RTs at the shortest distance (1 clause), faster RTs at the medium distances (5 and 10 clauses), and slower times again at the long distances (15 and 20 clauses). This main effect of referential distance was statistically significant ($F(5,165) = 4.16$, $p<.002$).

If pronouns are more efficient grammatical devices for signalling coreference over short distances, then one might have expected the 1-clause distance to produce a **short** rather than a long probe RT. This would have resulted in a monotonic increase in RT with increasing referential distance, rather than the U-shaped pattern which we observed. The unexpectedly high RT at the shortest referential distance may be a result of the probe's position early in the story. This explanation is supported by an analysis of the RTs for control probes presented early, in the middle or late in the passages. This analysis revealed longer RTs for probes presented early in the story frames (663 msec.), and shorter RTs for probes presented in the middle (615 msec.) or late (606 msec.) parts of frames.

**(6) Sample test passage for Experiment 3**

FRAME OPENING:
The restaurant was just about full when THE MAN arrived. Everybody was dashing about madly, too busy to notice him.

FRAME ENDING:
*He waited patiently for his turn.

1-CLAUSE REFERENTIAL DISTANCE:
[Text as above (opening + ending)]

LONGER REFERENTIAL DISTANCES:
[text inserted, cumulatively, before ending]

5-CLAUSE:
It was late on a Friday night, a time favored by the after-show crowd. All the tables were packed with hungry patrons. The waiting line threaded around like a coiled snake, and the noise from the kitchen was deafening...

10-CLAUSE:
The temperature in the main dining room must have reached ninety. It was late in the summer, and the air conditioning seemed to have given out. The faces in the waiting line were glistening with sweat, and many of the ladies had pulled their skirts up...

15-CLAUSE:
The lights were dimmed discreetly, coming from behind thick matted glass panels near the ceiling. The decor was basic continental with theatrical posters, live flowers and checkered table clothes...

20-CLAUSE:
People kept coming in and joining the line, some in pairs, others in large groups, and a few alone. The lobby was getting more and more crowded. The line twisted back over itself and eventually spilled out into the street...

The five different versions of the stories were divided into three stimulus lists. Each list contained two versions of each story and an equal number of paragraphs of each length. Each subject saw only one list, so that within and across subjects each referential distance was tested an equal number of times. Thus, this experiment had one repeated-measure variable (referential distance of 1,5,10,15 or 20 clauses).

The experimental probes were presented with the anaphoric pronouns at the end of each test frame. Control probes were randomly distributed throughout the story, but never appeared within 10 words of the probed pronoun. The number of control probes varied from one in the shorter versions of the frames to three in the longer versions. Control probes appeared at the same locations for all subjects tested with the same list.

A sheet consisting of 20 true/false comprehension questions, one were introduced as indefinite or (thematic) definite. Nonetheless, they were treated differentially in a way that seems to be governed solely by their **grammatical form:** *when such a participant was* coded with the indefinite article "a", the longest keypress time occurred on the noun itself [n], just as it did for pronouns (Experiment 1). On the other hand, when the same character was introduced with the definite article "the", the longest keypress time occurred on the word **following** the probed item, [n+1], as it does in the case of anaphoric definites, which are also marked by "the".

Having identified in the first two experiments what seems to be a grammar-triggered effect on the press times, at least in the presence of probes, we return in the next two experiments to the effect of referential distance on the processing of pronouns and anaphoric-definite nouns.

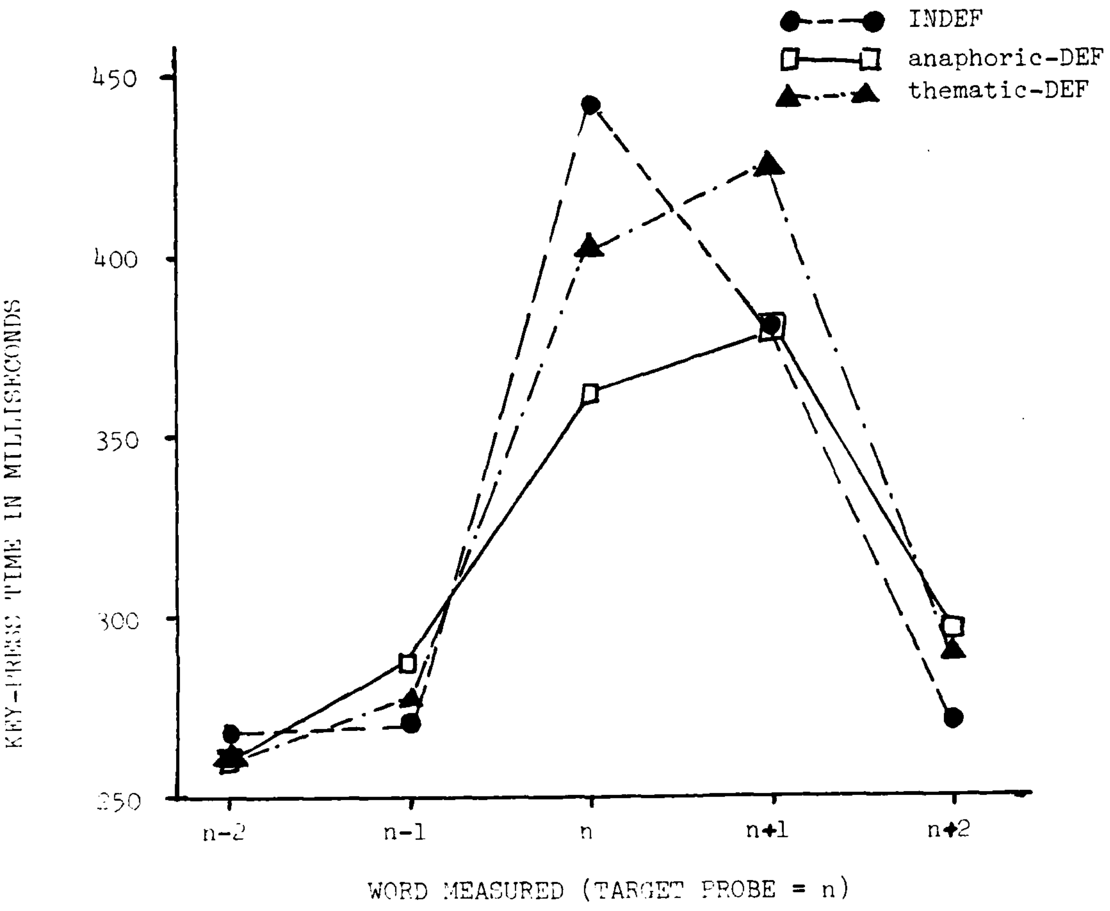
## 3.3. Experiment 3

### 3.3.1. Method

#### 3.3.1.1. Materials and design

The story frames in this experiment involved a single character and anaphoric reference employing only pronouns. Ten different stereotyped story frames were used, each with five variants that differed from each other primarily in the referential distance between the last prior mention of the participant and the subsequent reference to it by a pronoun. The filler clauses between the prior mention and the subsequent pronominal reference introduced only general descriptions of the scene and surrounding circumstances. These descriptions were all highly stereotyped and deliberately stayed away from the main character itself or his/her point of view. The referential distances tested were 1,5,10,15 and 20 clauses. The last sentence in the text always began with the coreferent pronoun, itself the locus of the probe, as in previous experiments. A sample text frame for this experiment is given in (6) below.

FIGURE III: Mean Keypress Times for Indefinite, Anaphoric-DEF and Thematic-DEF Nouns



Statistical analysis of these results with both referential device (INDEF, anaphoric-DEF, thematic-DEF) and word position (n-5,...,n,...,n+5) as repeated measures revealed a main effect of word position ($\underline{F}(10,230) = 22.28$, $p<.001$) due to the general increase in keypress times around probed items. In addition, however, there was a significant interaction between referential device and word position ($\underline{F}(30,690) = 1.79$, $p<.007$), indicating that the position of the peak keypress time varied depending on the type of referential device.

The results clearly rule out the possibility that the difference between DEF-nouns and pronouns reported in Experiment 1 is due to the difference in phonological/graphemic length between nouns and pronouns. What is more, the thematically stereotyped second participants in the stories were equivalent in terms of their predictability from the thematic structure of the frame, as well as in terms of their "new information" status, regardless of whether they

### 3.2.1.3. Procedure

The same procedure was used in this experiment as in Experiment 1.

### 3.2.2. Results

### 3.2.2.1. Comprehension test

The subjects' mean percent correct on the comprehension test was 88%.

### 3.2.2.2. Probe reaction time

The means of the subjects' median reaction times to the visual probes in each condition are given in Table III, below. The RTs are fastest for the anaphoric DEF-nouns and slowest for the INDEF-nouns, but none of the differences are statistically significant.

TABLE IV: Median Probe Reaction Time for Indefinite, Anaphoric-DEF, and Thematic-DEF Nouns

| CATEGORY | RT |
|---|---|
| INDEF | 624 |
| Anaphoric-DEF | 605 |
| Thematic-DEF | 613 |
| Control probe | 622 |

### 3.2.2.3. Keypress times for probed words

The mean keypress times (in msecs) across subjects for anaphoric DEF-nouns, thematic DEF-nouns and INDEF-nouns, as well as for the two words directly preceding and following them in the text, are given in Table IV, below, and represented graphically in Figure III, below.

TABLE IV: Mean Keypress Times for Indefinite, Anaphoric-DEF and Thematic-DEF Nouns

| CATEGORY | n-2 | n-1 | n | n+1 | n+2 |
|---|---|---|---|---|---|
| INDEF | 267 | 271 | 438 | 383 | 275 |
| ANAPHORIC-DEF | 264 | 288 | 364 | 382 | 298 |
| THEMATIC-DEF | 268 | 277 | 404 | 423 | 290 |

was described entering into a scene in the first sentence. The next 1-3 sentences described the character's progress through the scene in a rather stereotyped fashion. The final sentence then began with one of the following three referential devices:

(a) Anaphoric-DEF noun: The (identical) noun which introduced the single character at the beginning of the frame;

(b) Thematic-DEF noun: A second character, highly predictable from the thematics of the frame, introduced as a DEF-noun (thus having the same grammatical form as (a);

(c) Indefinite noun: The same second character as in (b), but here introduced as an indefinite noun.

Each of the twelve experimental story frames used in this experiment appeared equally often across subjects in all three conditions. A sample frame is given in (5) below.

(5) **Sample text passage for Experiment 2:**

"The restaurant was just about full when THE MAN arrived. Everybody was dashing about too busy to notice him...

(a) Anaphoric- DEF ending: ...The *man waited patiently in the corner".
(b) Thematic-DEF ending: ...The *waitress came over finally and gave him the menu".
(c) INDEF ending: ...A *waitress came over finally and gave him the menu".

In addition to the twelve experimental text frames, ten filler texts of equal length (but different in structure) were also given to the subjects.

The different versions of each story were divided into two lists so that two versions of each story were on a list. Each list was split into two blocks so that one version of each story was in each block. Five stories in each block were filler paragraphs. Subjects were randomly assigned to one of two groups, each group being tested with only one list. Half of the subjects in each group were tested with one block of the list first and the other half were tested with the other block first.

Experimental probes were presented with the nouns of the referential noun phrases in the last sentence of each paragraph. Half of the paragraphs had an additional control probe presented at a pseudorandomly chosen location early in the paragraph. Booklets with a true/false comprehension question corresponding to each paragraph in a list were prepared.

### 3.2.1.2. Subjects

Twenty-four subjects were recruited from the same subject pool used in Experiment 1, with the same provisos as to their native language, vision, reading status and pay. None of the subjects had participated in Experiment 1.

TABLE II: Mean Keypress Times for PRO and DEF-Nouns
at 1 and 5 Clause Referential Distance

| | Word Position | | | | |
|---|---|---|---|---|---|
| Condition | n-2 | n-1 | n | n+1 | n+2 |
| PRO 1-cl | 288 | 295 | 437 | 320 | 261 |
| PRO 5-cl | 288 | 298 | 377 | 345 | 303 |
| DEF 1-cl | 284 | 313 | 344 | 361 | 259 |
| DEF 5-cl | 293 | 318 | 309 | 383 | 259 |
| Control probe | 307 | 306 | 354 | 375 | 284 |

Probe reaction times for control items and the keypress times around these items were also analyzed. The mean of subjects' median probe reaction times was 620 msecs (Table I), longer than the RTs observed in any condition except pronouns at the 5-clause referential distance. The difference was reliable for DEF-nouns ($\underline{F}(1,11) = 21.34$, $p<.001$), but not for pronouns ($\underline{F}<1$). The control probe RT was significantly longer than the RT for probed anaphoric devices at the 1-clause referential distance ($\underline{F}(1,11) = 11.65$, $p<.006$), but not at the 5-clause referential distance ($\underline{F}<1$). The relatively long probe RTs for control items may reflect their earlier positions in the passages compared to anaphoric probed items. Since the controls constitute a diverse set of lexical items, we are hesitant to interpret absolute differences between these items and the anaphoric items under study. The pattern of response times to control items, however, has proved useful in discerning general effects in both the probe RTs and keypress times.

The longest keypress time for the control probes was on the word following the probed item [n+1]. This is the pattern that was found for DEF-nouns. However, when the keypress times for the probed item and the word preceding and following it is compared across pronouns, DEF-nouns and the control items, no significant differences are found.

It might be argued that the significant one-word shift in the location of the peak keypress time for probed DEF-nouns (as compared with probed pronouns) may simply be due to the fact that nouns are phonologically (here graphemically) longer than pronouns, and thus take longer to encode before the processing of coreference occurs. The next experiment was designed to investigate such a possibility.

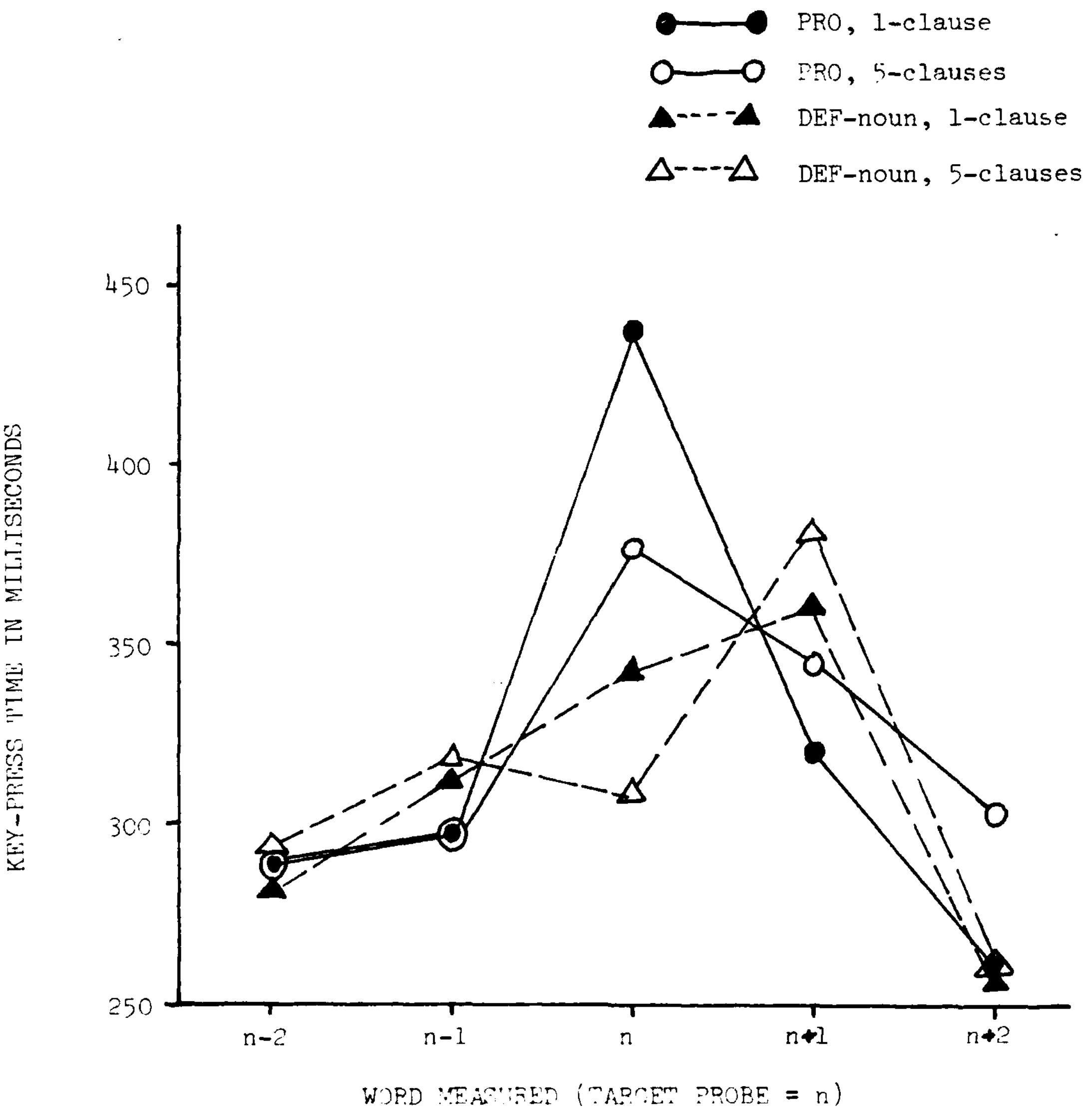
## 3.2. Experiment 2

### 3.2.1. Method

### 3.2.1.1. Materials and design

The story frames for this experiment were rather short and involved no manipulation of referential distance. A single character

FIGURE II: Mean Keypress Times for Pronouns and Anaphoric Definite Nouns at 1 and 5 Clause Referential Distance

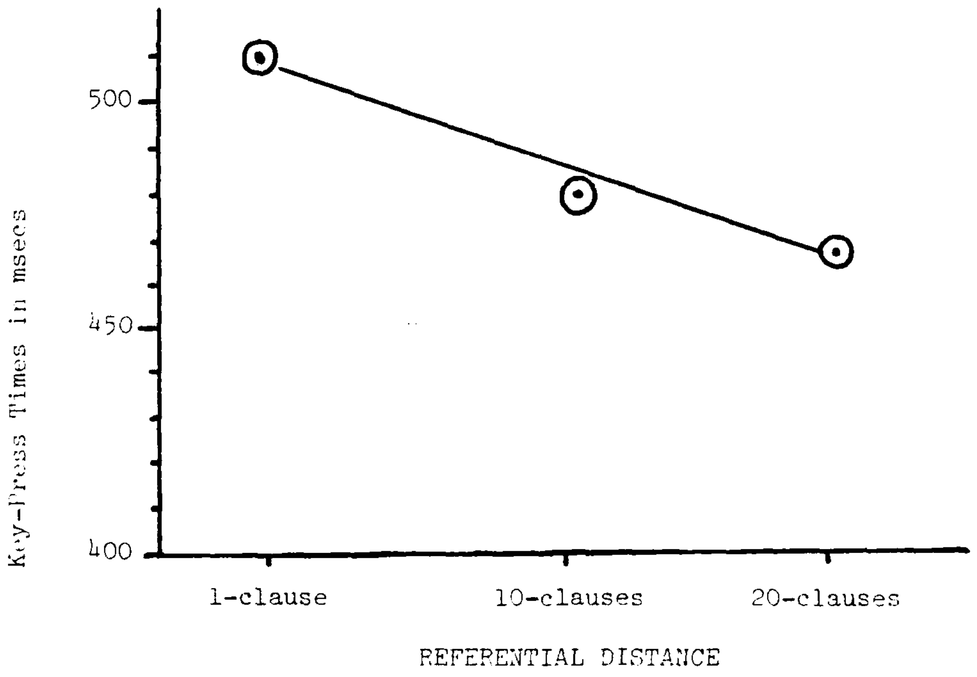


anaphoric-DEFs contrasts with the keypress times for control items at [n+1], which show the same U-shaped pattern we observed in the probe RTs.

For thematic-DEF nouns the pattern in this experiment is less consistent. For the 10-clause distance, the peak keypress time appears on [n+1] as in Experiment 2, but at the 1 and 20 clause distances, the peak keypress time occurs on [n].

Overall, the keypress times collapsed across word position and type of referential device were FASTEST at the 10-clause referential distance, SLOWEST at the 1-clause distance, and intermediate at the 20-clause distance (thus exhibiting the U-shaped curve shown in the control-probe results). Collapsed across referential distance and word position, keypress times were SLOWER for thematic-DEF than for anaphoric-DEF nouns.

FIGURE: VI Mean Keypress Times for Anaphoric DEF-Nouns at the [n+1] Word Position at 1, 10 and 20-Clause Referential Distances



A 2 (type of referential device) by 3 (referential distance) by 11 (word position) repeated measures analysis of variance of the mean keypress times for the probed words and the 5 words directly preceding and following them revealed a significant main effect of referential distance ($F(2,20) = 3.26$, $p < .05$), with mean RTs of 411, 376 and 393 msecs., respectively, for the 1, 10 and 20 clause distances. Main effects of referential device ($F(1,35) = 7.5$, $p < .01$), reflecting the longer RTs for thematic-DEFs than for anaphoric-DEFs, and of word position ($F(10,350) = 16.54$, $p < .01$), reflecting the generally increased RTs around probed items, were also found. There was a marginally significant interaction of referential device and referential distance ($F(2,70) = 2.65$, $p < .08$), suggesting that thematic-DEFs were slower than anaphoric-DEFs only at the 1-clause distance (70, 9 and 2 msecs. difference at each distance, respectively). Significant interactions between referential distance and word position ($F(20,700) = 2.05$, $p < .005$), as well as between referential device and word position ($F(10,350) = 2.20$, $p < .02$) were found. The interaction between referential device and word position reflects the peak keypress time for anaphoric-DEFs at [n+1], and the tendency of the peak keypress time for themtic-DEFs in this experiment to be at [n]. However, this interaction was qualified by the three-way interaction between referential device, referential distance and word position ($F(20,700) = 1.92$, $p < .01$), reflecting the fact that the peak keypress time at the 10-clause distance was at [n+1] for both anaphoric- and thematic-DEFs.

## 3.5. Experiment 5

### 3.5.3. Method

### 3.5.1. Materials and design

In this experiment we compared two referential devices (anaphoric DEF-nouns and pronouns) at two referential distances (1 and 12 clauses). Ten different story frames were used, each introducing two characters (one male and one female) early in the discourse. As in Experiment 1, the next segment of the story continued to deal with the second character only. The final sentence then referred to either of the two, either by pronoun or anaphoric-DEF noun. A sample text passage can be seen in (8) below.

(8) **Sample test passage for Experiment 5**

The woman entered the waiting room trailed by her five-year-old boy. He immediately proceeded to jump on the couch and bounce on it energetically. It felt nice and soft. He bounced higher and higher, looking about him in glee. It was going to be quite an adventure, he could tell, with all these new things around him; such excitement with objects to take apart. What sheer fun! He turned his attention to the window, surveying through its wide pane the parking lot three floors below. Now here was a real challenge. A ledge he could climb onto. An adventure waiting for him right there...

1-CLAUSE DISTANCE ENDINGS:

PRONOUN: *He clambered onto the windowsill and fiddled with the latch.
DEF-NOUN: The *boy clambered onto the windowsill and fiddled with the latch.

12-CLAUSE DISTANCE ENDINGS:

PRONOUN: *She stepped over, leaned forward and said something in a hushed voice.
DEF-NOUN: The *woman stepped over, leaned forward and said something in a hushed voice.

In this experiment we controlled for the amount of information presented in a trial before an item was probed, thus the passages were all of equal length. There were four versions of each story, divided into two stimulus lists. Each list contained an equal number of passages in each condition. Each subject was given only one of the lists. Thus, stories appeared an equal number of times in each condition across subjects. As in the previous experiments, true/false comprehension questions were prepared for each story.

The test probes were presented either with the pronoun or anaphoric-DEF noun in the last sentence of each passage. Two to three control probes were also inserted in each passage, with the restriction that they not appear any closer than (approx.) 20 words from the probed anaphoric item.

### 3.5.1.2. Subjects

Eighteen subjects were recruited from the same pool as described above, subject to the same conditions outlined above.

### 3.5.1.3. Procedure

The procedure used in this experiment was the same as in Experiment 4, above.

### 3.5.2. Results

#### 3.5.2.1. Comprehension test

The subjects' mean percent correct on the comprehension test was 90%.

#### 3.5.2.2. Probe reaction time

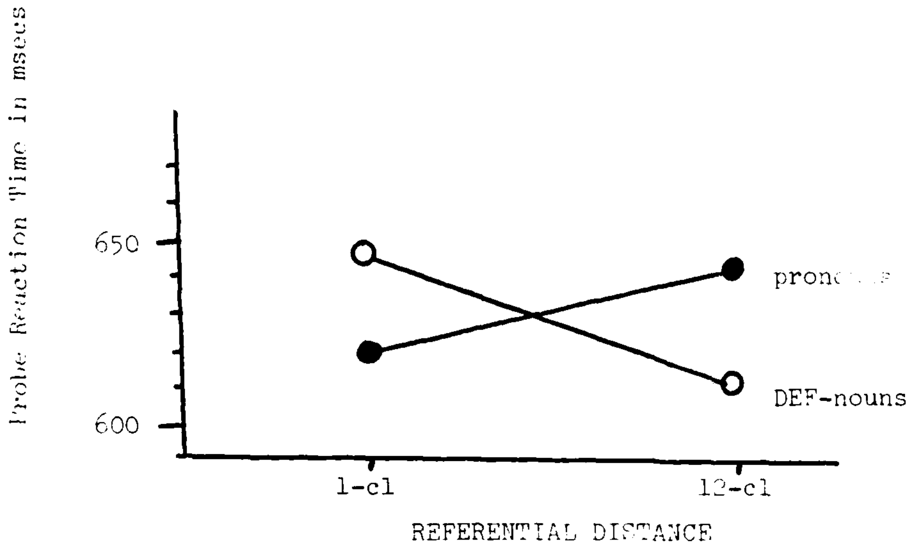
The median probe reaction time for pronouns and DEF-nouns at the 1-clause and 12-clause referential distances are given in Table IX, below.

TABLE IX: Median Probe Reaction Time for Pronouns and Anaphoric DEF-nouns at 1 and 12 Clause Referential Distance

| | REFERENTIAL DISTANCE | |
|---|---|---|
| CONDITION | 1-CLAUSE | 12-CLAUSES |
| Pronoun | 619 | 638 |
| DEF-noun | 645 | 607 |

The results are expressed graphically in Figure VII, below. For DEF-nouns, probe RT is faster at the 12-clause distance than at the 1-clause distance. For pronouns, probe RT is slower at the 12-clause distance than the 1-clause distance. These results replicate those obtained for 1 and 5 clause distances in Experiment 1 (see Figure I). Statistical analysis of the probe RT results, however, indicates that none of the differences reached significance ($F$'s < 1 for the effects of referential distance and referential device, and for their interaction $F(1,17) = 2.40$, $p < .14$).

FIGURE VII: Median Probe Reaction Time for Pronouns and Anaphoric-DEF Nouns at 1 and 12 Clause Referential Distance



### 3.5.2.3. Keypress times for probed words

The results of the keypress times for the various conditions are given in Table 4, below. As before, probed pronouns show the peak keypress time on the probed item itself [n], and the probe effect on reading times is, as before, much more pronounced than with anaphoric-DEF nouns. In this experiment, however, the referential distance effect for pronouns was not obtained; that is, the peak keypress time at [n] was not higher at the 12-clause distance than at the 1-clause distance.

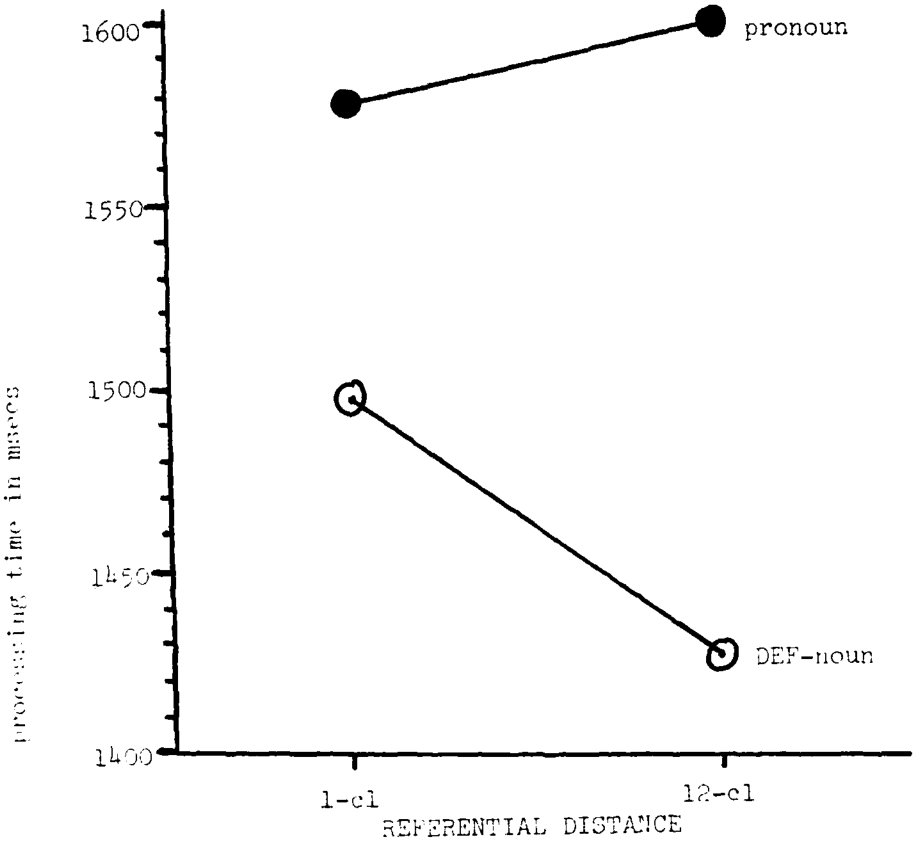
For DEF-nouns, the results are somewhat mixed. We have duplicated here the results of the relevant previous experiments concerning the location of the peak keypress time for the 1-clause distance, but not for the 12-clause distance, where the press time peak appears instead on the probed item. Also duplicated are the smaller keypress values for DEF-nouns than for pronouns. Finally, when the keypress times of both [n] and [n+1] are combined, the total processing time at the 12-clause referential distance is shorter than at the 1-clause distance.

The results of this experiment may also be expressed by combining the probe RT and keypress time measures at [n] and [n+1], for pronouns and DEF-nouns at each referential distance. This combined expression is plotted graphically in Figure VIII, below.

TABLE X: Mean Keypress Times for Pronouns and Anaphoric-DEF Nouns at 1 and 12 Clause Referential Distance

| | WORD POSITION | | | | |
|---|---|---|---|---|---|
| CONDITION | n-2 | n-1 | n | n+1 | n+2 |
| PRONOUN: | | | | | |
| 1-clause | 309 | 403 | 560 | 421 | 290 |
| 12-clause | 316 | 307 | 565 | 380 | 338 |
| DEF-NOUN: | | | | | |
| 1-clause | 321 | 350 | 442 | 471 | 344 |
| 12-clause | 337 | 328 | 446 | 372 | 372 |

FIGURE VIII: Probe RT + Keypress Time Combined for Positions [n] and [n+1] for Pronouns and DEF-Nouns at 1 and 12 Clause Distance



Statistical analysis of the results reveals a significant main effect of word position ($\underline{F}(10,30) = 5.74$, $p<.001$). The interaction between referential device and word position is marginally significant ($F(10,30) = 1.80$, $p<.08$), reflecting the tendency for pronouns and anaphoric-DEF nouns to have peak keypress times at different locations ([n] vs. [n+1], respectively). The interaction between referential distance (1 vs. 12 clause) and word position is significant ($\underline{F}(10,30) = 2.70$, $p<.01$).

### 3.6. Experiment 6

This experiment was designed for two purposes. First, to test by an independent method whether subjects in fact re-instantiate the referent when they encounter a pronoun or DEF-noun as a referential device (see Corbett & Chang, 1983). And second, to compare by one more method the efficacy of pronouns vs. DEF-nouns as coreferential devices at different referential distances.

### 3.6.1. Method

### 3.6.1.1. Materials and design

There were 18 basic story frames in this experiment, all involving a single character. After introducing the character in the first clause, five clauses followed describing the scene but leaving the character out. The next sentence then reinstated the character with either a pronoun, or the identical DEF-noun used to introduce it. The probes in this experiment always occurred at the location of these reinstating devices, either in lieu of the device itself or immediately after it. In half of the stories, the introduction of the single character was preceded by five clauses of descriptive material in order to vary the structure of the passages.

Rather than the flashing light probe used in previous experiments, in this experiment the probe was a word, appearing at the top center of the screen. Upon presentation of a probe, subjects were told to decide as quickly as possible whether the word had appeared in the story and to indicate their decision on a two-key response board. Correct ('yes') probe words were the nouns used to introduce characters initially. Incorrect ('no') probe words were nouns used to introduce a character in other stories in the stimulus list. The two probe types appeared with equal frequency in a list.

The probe words were presented under three experimental conditions:

(a) Immediately after coreference (reinstatement) by a propnoun;
(b) Immediately after coreference (reinstatement) by a DEF-noun;
(c) In place of the word that would have reinstated the character (no reinstatement).

A sample test passage can be seen in (9) below.

(9) **Sample test passage for Experiment 6:**

The PLUMBER brought most of his small tools into the house before beginning to install the new set of pipes. The shop had received a call concerning the problem early that morning. The garbage disposal was being replaced when the wrench slipped, causing the pipe to crack open and spew out bits of stale water and corroded metal. The house was at least seventy years old with historic, as well as sentimental, value....

ENDINGS:
(a) Pronoun: He*...
(b) DEF-noun: The plumber*...
(c) Control: *...

PROBE WORDS:
(a) 'yes' probe: PLUMBER
(b) 'no' probe: PRIEST

Three stimulus lists were prepared, consisting of two versions of each story frame (a 'yes' and a 'no' probe word trial). A list was presented to a subject in two blocks. Each subject saw only one list. Subjects were assigned to three groups, each group tested with a different list. Half of the subjects within a group were tested with one block first, the other half saw the other block first. Across subjects, each test passage appeared an equal number of times in each condition. A true/false comprehension test for each story was administered, as in the preceding experiments.

### 3.6.1.2. Subjects

Thirty subjects were recruited from the same pool as described above, subject to the same conditions as outlined above.

### 3.6.1.3. Procedure

The presentation of test passages followed the same procedure as in Experiment 3, above, except that once the probe word had been presented and the response obtained, the trial was discontinued.

Subjects were told to press the key on the RIGHT of the response board if the probe word had appeared in the passage ('yes' trial), and the key on the LEFT if the probe word had not appeared in the passage ('no' trial). After the probe reponse had been made, the screen went blank and the subject answered a true/false comprehension question about the passage they had just read. Probe RTs and reading times were measured and recorded by an Apple IIe computer, as above.

### 3.6.2. Results

### 3.6.2.1. Comprehension test

The subjects' mean percent correct on the comprehension test was 86%.

### 3.6.2.2. Probe reaction time

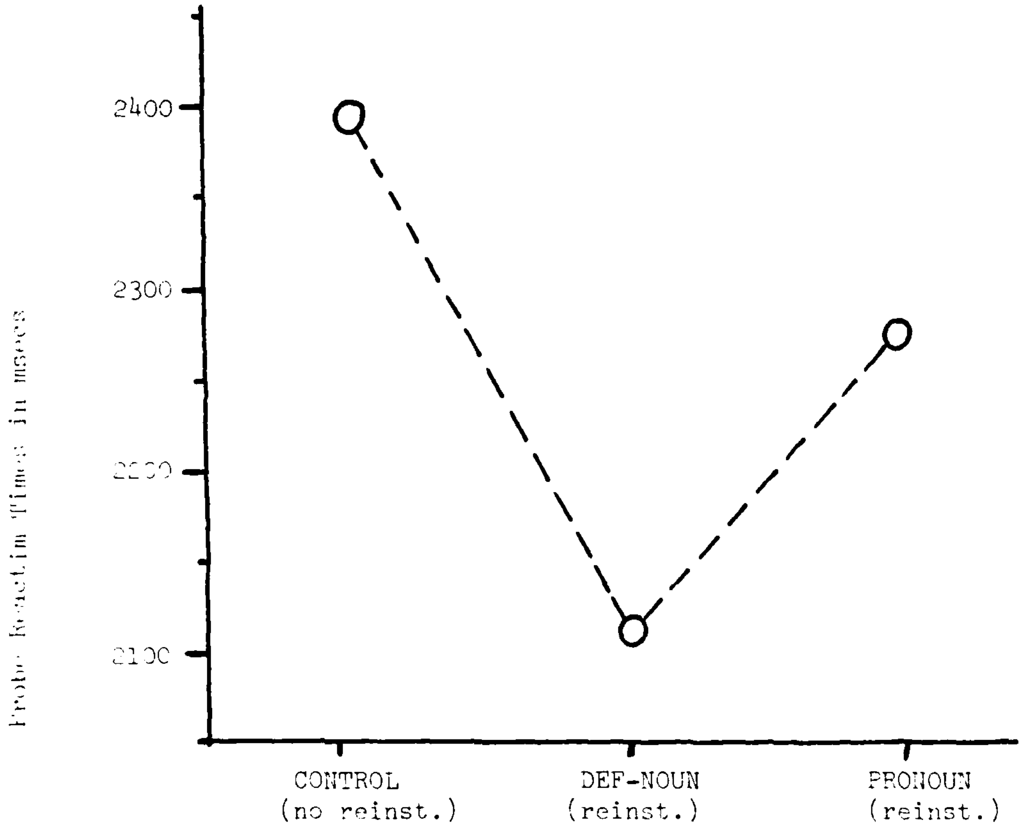
Subjects' median response times (in msecs) to the 'yes' and 'no' probe words in the three conditions (pronoun reinstatement, DEF-noun reinstatement and control (no reinstatement)) are given in Table XI, below. The 'yes' probe results are plotted in Figure IX, below.

TABLE XI: Median Probe Reaction Time for Character Reinstatement with Pronoun or DEF-noun and for Control (No Reinstatement) Condition

| | CONDITION | | |
|---|---|---|---|
| PROBE RESPONSE | PRONOUN | DEF-NOUN | NO REINSTATEMENT |
| 'yes' probe | 2276 | 2009 | 2393 |
| 'no' probe | 2092 | 2306 | 2211 |

The control ('no reinstatement') condition displays the slowest RTs for 'yes' probe words. Reinstatement with an identical DEF-noun results in faster probe verification times than reinstatement with a pronoun. Statistical analysis of the 'yes' responses only reveals a significant main effect of reinstatement condition (*F*(2,46) = 3.59, p .04). More specifically, the difference between the DEF-noun reinstatement and the no reinstatement condition is significant (*F*(1,23) = 7.09, $p<.02$). However, the results for the pronoun reinstatement condition are equivocal: the difference between the pronoun reinstatement and no reinstatement condition is not significant (*F*<1), but neither is the difference between the pronoun and DEF-noun reinstatement condition (*F*(1,23) = 2.56, p .12).

FIGURE IX: Median Probe Reaction Time to "Yes" Probe Words for Pronouns, DEF-Nouns and No-Reinstatement Control



## 4. Discussion

The results of our six experiments, despite occasional variability and lower than desired statistical significance, point to two seemingly replicable correlations between the **grammatical form** of referential devices, and the **way** subjects process such devices in written discourse. We discuss the two effects separately, but make some tentative suggestions in section 4.3 as to how they might be related.

### 4.1. The word position effect

We have found that when the rate of reading is measured, and a visual probe is introduced simultaneously with a referential noun or pronoun, the reading rate is momentarily slowed with the maximum pause (peak press times) at different points, depending on the grammatical ce. is used:

(a) **DEF-nouns**, either **anaphoric** ones which refer to a participant explicitly introduced earlier in the discourse, or **thematic** ones which introduce a new participant into the discourse, display a peak press time on the word following the probed item [n+1]. As of now, this pattern is indistinguishable from the control pattern.

(b) **INDEF-nouns**, which introduce new participants into the discourse, and **pronouns**, which refer to a participant explicitly introduced earlier in the discourse, display a peak press time on the probed word itself [n]. This pattern is sharply distinguished from the control pattern.

Since definite and indefinite nouns are not grouped together, and since pronouns and indefinite nouns are grouped together, the word position effect could not merely be due to the lexical/phonological difference between nouns and pronouns. Rather, one must seek to explain the difference in terms of the different functions these various grammatical devices perform in the processing of referents in discourse.

We would like to suggest, rather tentatively at this point, that definite nouns trigger a (grammar-signalled) processing of the referent which makes the most demands on cognitive resources at the point in time corresponding to the word following the referential noun. This delay is necessary in order for the reader/hearer to make a preliminary determination as to the **"file source" of definiteness** involved. Linguists have noted for a long time (see summary in Givon, 1984, ch. 11) that definiteness -- i.e. the assumption that the referent is accessible to the hearer -- can arise from a large variety of possible file sources. To illustrate this, consider the following examples:

11 **File sources of definite nouns in discourse:**

a. I got up this morning and **the sun** was shining...
FILE: Generic knowledge of the physical universe.
SHARED BY: All members of human culture.

b. ...So I turned **the TV** on and there was **the President**...
FILE: Generic knowledge of culture.
SHARED BY: All members of the culture.

c. ...So I sat on the chair, but **the seat** was broken...
FILE: Generic-cultural knowledge of wholes & parts of objects.
SHARED BY: All members of the culture.

d. John's **father** was an honest man
FILE: Generic bio-cultural knowledge.
SHARED BY: All members of the culture.

e. Take **this chair** and put it under **that table**
FILE: Deictic knowledge of the immediate speech situation.
SHARED BY: Specific speaker and hearer **in situ**.

f. We went to the bank, and **the cashier** cashed my check
FILE: Generic-cultural knowledge of normative scripts/scene
("**thematic** definite").
SHARED BY: All members of culture/sub-culture.

g. We saw a man and a woman. **The woman** was tall and...
FILE: Specific discourse file ("anaphoric definite").
SHARED BY: Specific speaker and hearer.

The convention of shared definite reference, in terms of the basis for the speaker's assumption that the hearer can uniquely identify a referent, thus arises from three major contextual files or **foci**:

Generic file: culturally-shared focus;
Deictic file: situationally-shared focus;
Textual file: discourse-shared focus.

The processing delay triggered by grammatical marking with the DEF article occurs on the item following the referential word. If the peak press time reflects the achievement of coreference, why should it be located on the word following the referential device for DEF nouns but on the word itself for INDEF nouns and pronouns? Our suggestion is that the peak delay at [n+1] for DEF nouns may reflect the **preliminary decision** as to the subsection of the definite filing area that should be searched for the referent. Once that decision has been made, retrieval of the referent from the appropriate file and the achievement of coreference can proceed. In contrast, INDEF nouns require no such decision, since no file is relevant, and the reader/hearer can proceed immediately to open a new file.

Finally, the reason anaphoric pronouns produce an immediate delay must involve the fact that unlike DEF nouns, their co-reference source is extremely uniform in human discourse. To illustrate this, consider the following distributional data, taken from a quantified text-based study of spoken English narrative (Givón, 1983b). In this study, the referential distances of zero anaphora, pronouns and various types of definite noun phrases were compared. In the results, given in Table XII, below, we have grouped together the data of pronouns and zero anaphora. This is justified by two considerations: first, zero anaphora is a relatively infrequent category in the English text as compared with anaphoric pronouns, and second, cross-language studies (Givón, 1983a) clearly show that the average referential distance for zero anaphora and pronouns is virtually identical (both means between 1.0 and 1.5, both medians 1.0). We contrast the values of the combined pronoun category with that of the combined definite nouns.

TABLE XII: The distribution of referential distance for pronouns (including zero anaphora) and definite nouns in spoken English narrative (from Givón, 1983b)

| | pronouns | | DEF-nouns | |
|---|---|---|---|---|
| # of clauses | n | % | n | % |
| 1-2 | 499 | 92.2 | 54 | 26.7 |
| 3-9 | 31 | 5.7 | 36 | 17.8 |
| 10-14 | [illegible] | 1.4 | 17 | 8.4 |
| 15-19 | 2 | 0.3 | 6 | 2.9 |
| 20+ | 1 | 0.1 | 89 | 43.9 |
| total: | 541 | 100.0 | 202 | 100.0 |
| median: | 1.0 clause | | 12.0 clauses | |

These distributional data are rather striking. When encountering an anaphoric pronoun in discourse, the probability is 99.6% that its coreference can be recovered by scanning the preceding 19 clauses in the discourse file. The probability is 92.2% that the coreference can be recovered by scanning only the **last two clauses** of the preceding discourse. In human language, such a distribution is virtually categorical.

In sharp contrast, when encountering a DEF-noun, the probability is 43.9 percent that no antecedent will be located within the preceding 19 clauses in the discourse. The scanning of the immediately preceding two clauses in the discourse could yield recovery of the referent in only 26.7% of the cases.

### 4.2. The referential distance effect

The second major result we have obtained, again with some degree of variability but nonetheless considerable replication, is the different effect of referential distance on the processing time for pronouns and definite nouns. Briefly, the efficacy of definite nouns for indicating co-reference is LOW at the minimal referential distance of 1 clause, but INCREASES toward 5-10 or 15-20 clauses. Exactly the opposite is true for pronouns. The distribution data in Table XII above provide a natural explanation for this difference. Pronouns find their antecedents in the **immediately preceding** discourse environment -- 86.2% within the preceding 1-2 clauses. As grammatical clues in the search for coreference, their efficacy is thus highest in that range, and decreases with the increase in referential distance. In contrast, only 26.7% of definite nouns find their antecedents within that short range. Another 17.8% find their antecedents within the preceding 3-8 clauses; and a full 43.9% find no coreferent within the preceding 20 clauses. As grammatical clues, DEF-nouns are thus much less efficient at the 1-clause minimal range, and their efficacy should increase at longer ranges.

We find, then, that referential distance by itself -- and thus memory decay -- does not suffice as a predictor of processing time. Rather, the difficulty of achieving coreference is mediated by grammar.

### 4.3. Automated vs. attended processes in the tracking of referents

As suggested in our introduction (section 1.3., above), grammar can be viewed as a routinized processing mode, whereby highly conventionalized morpho-syntactic clues trigger automatic processing responses. We would like to suggest now, again rather tentatively, that our results can be interpreted as an interaction between automated/unattended and analytic/attended components in the processing of referents in discourse. In this system, the grammatical coding of a referent as either pronoun, DEF-noun or INDEF-noun triggers some automatic responses. However, in addition, other non-automated, attended, context-scanning processes must be integrated into the overall process of the tracking of referents. This overall process may be illustrated as a flow chart, as in Figure 4, below.

FIGURE Y: Flow Diagram for Decision Processes
in Referential Tracking

| GRAMMAR CLUE | RAPID DECISION PROCESSES | ATTENTIONAL PROCESSES |
|---|---|---|
| pronoun | activate co-reference scan ===> | scan 1-2 clauses in immediately preceding discourse |
| INDEF-noun | activate opening of new file ===> | open new file |
| DEF-noun | defer activation of scanning and:<br>determine SOURCE of definite file; then: | |
| | (i) If anaphoric ===> | scan preceding 10-15 clauses in discourse for coreference |
| | (ii) If thematic ===> | scan discourse frame for likely thematic trigger |
| | (iii) If cultural ===> | scan general culture file appropriate to the particular item in the discourse |
| | (iv) If situational ===> | scan the immediate face-to-face situation |

We thus suggest that there is a grammar-triggered response that takes place initially, making processing decisions such as:

(a) Triggering a highly localized scan of the preceding discourse for the antecedent of an anaphoric PRONOUN.
(b) Opening a new file for an INDEFINITE referent;
(c) Delaying the in-file search for the antecedent of a DEFINITE referent, in order to first determine the **file source** responsible for the definite reference;

On the other hand, the scanning process itself -- searching inside the appropriate file and identifying the right antecedent -- is a non-automated, attention-demanding process. This is shown in two ways in our experiments. First, attending to a secondary probe slows down that process; and second, this process -- both in terms of probe RT and the rate of reading in the presence of a probe -- is sensitive to increased referential distance in highly specific ways.

The effect of grammatical routinization also manifests itself in the differential sensitivity of pronouns and DEF-nouns to increased referential distance. Pronouns are grammatical clues that trigger an efficient short-range coreference search, through the immediately preceding discourse. When the correct antecedent in fact occurs outside of this range, a cost in processing time is incurred. DEF-nouns, once it has been determined that they arise from the anaphoric text-source file, trigger a more long-range search.

## REFERENCES

Aaronson, D. & Ferres, S. (1983). A model for coding lexical categories during reading. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 700-725.

Anderson, A., S.C. Garrod and A.J. Sanford (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text, *Quarterly J. of Experimental Psychology*, *35A*.

Corbett, A.T. and F.R. Chang (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition*, *11*, 283-294.

Derbyshire, D. (1985). Topic continuity and OVS order in Hixkaryana. In J. Sherzer and G. Urban (Eds.), *Native South American Discourse*.

Fletcher, C.R. (1982). Markedness and topic continuity in discourse. *Journal of Verbal Learning and Verbal Behavior*, *23*, 487-493.

Givón, T. (1979). *On understanding grammar*. New York: Academic Press.

Givón, T. (1982). Logic vs. pragmatics, with human language as the referee: Towards an empirically viable epistemology. *Journal of Pragmatics*, *6*.

Givón, T. (Ed.) (1983a). *Topic continuity in discourse: Quantitative cross-language studies*. *TSL, Volume 3*. Amsterdam: J. Benjamins.

Givón, T. (1983b). Topic continuity in spoken English. In T. Givón (Ed.), *Topic continuity in discourse: Quantitative cross-language studies*. *TSL, Volume 3*. Amsterdam: J. Benjamins.

Givón, T. (1984). *Syntax: A functional-typological introduction, Vol. 1*. Amsterdam: J. Benjamins.

Just, M.A. & Carpenter, P.A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329-354.

Keele, S. (1968). Motor control in skilled motor performance. *Psychological Bulletin*, *70*, 387-403.

Kennedy, A. & Murray, W.S. (1984). Inspection times for words in syntactically ambiguous sentences under three presentation conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 833-849.

Kerr, D. (1973). Processing demands during mental operations. *Memory & Cognition*, *1*, 401-412.

Payne, T. (1985). Participant coding in Yagua. Unpublished doctoral dissertation, University of California at Los Angeles.

Posner, M.I. & Marin, O.S.M. (Eds.). *Attention and Performance XI*, in press.

Posner, M.I. & Warren, R. (1972). Traces, concepts and conscious constructions. In A.W. Melton and E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: Winston and Sons.

Schneider, W. & Schiffrin, R.M. (1977). Controlled and automatic human information processing, I: Detection, search and attention. *Psychological Review*, *84*, 1-66.

Shapiro, D. & Schmidt, R. (1980). The schema theory: Recent evidence and developmental implications. In J.A.S. Kelso and J.E. Clark (Eds.), *The development of movement control and coordination*. New York: Wiley and Sons.

Tomlin, R. (1985). Linguistic reflections of cognitive episodes. In R. Tomlin (Ed.), *Coherence and grounding in discourse*, *TSL*, *Volume 11*. Amsterdam: J. Benjamins.

Whittaker, H. (1983). Towards a brain model of automatization: A short essay. In R. Magill (Ed.), *Memory and control of action*. Amsterdam: North Holland.

END

FILMED

9-85

DTIC